J. Fac. Edu. Saga Univ.
Vol.19, No. 2(2015) 127～139

Challenges and complexities of assessing the learning of a foreign language in primary schools: Insights from bilingualism and second language acquisition

Yuko HAYASHI※

ABSTRACT

Assessing language learning in children, let alone the learning of a foreign (FL) or second language (L2), poses a considerable challenge on teachers, researchers and policy makers. This paper provides a non-exhaustive overview of the issues and challenges surrounding language assessment from the perspectives of language acquisition and bilingualism, within which both relevant and unique challenges faced by FL pedagogy in Japanese primary school settings are described and illuminated further. Approaches to strengthening the interaction between assessment, pedagogy and curriculum designs are discussed towards the end.

Keywords: language assessment, bilingualism, language acquisition, foreign language learning

An increasing number of school-age children around the world are learning more than one language in school, be it a foreign (FL) or second (or third) language (L2), whilst at the same time learning and developing their first language (L1) by utilising various 'funds of knowledge' (i.e., skills, bodies of knowledge and cultural resources shared amongst their communities and the society) (Gregory, 2008). This global phenomenon is also reflected in the Japanese primary curriculum, where Foreign Language Activities (FLA), *Gaikokugokatsudō* in Japanese, was fully implemented nationwide as a compulsory area of study for pupils in Years 5−6 in 2011. The primary reasons for introducing FLA are as follows (MEXT 2008):

a) To meet the growing demand of the global society, where English is considered an essential tool for communication;
b) To foster intercultural sensitivity and communication abilities by drawing on the flexible adaptability of children;
c) To ensure equal educational opportunities afforded to children by distributing a set of teaching materials (entitled *Hi, friends!* 1, 2) to schools across the Nation and a smooth transition into secondary education.

In current practice, children attend a class of FLA once a week, totalling approximately 35 hours a year. As stipulated in the 2008 revised edition of the Course of Study for primary schools, the set of teaching guidelines issued by the Ministry of Culture, Education and Technology (MEXT, 2008), FLA is not taught as a subject such as mathematics and science and hence involves no formal, numerical evaluation of

※ Department of Education of School Subjects, Faculty of Culture and Education, Saga University

outcomes and progression in English language learning. It is envisaged, however, that the age at which children begin learning English as an FL (EFL) as part of the primary curriculum, which corresponds to 10−11 years as of today, will be lowered to 8−9 years in 2018, when the next edition of the Course of Study will be issued (MEXT, 2014). According to the MEXT (2014), FLA is likely to be introduced to pupils in Years 3−4, whereas English will be taught as a subject to pupils in Years 5−6. It follows then that there will be a formal evaluation of FL outcomes and progression, to which end, appropriate assessment will be carried out.

Language assessment is a challenge to teachers, researchers and policy makers and poses even greater challenges when it comes to assessing the learning of an FL/L2 in young learners, an issue the current study focuses on below. In so doing, the literature on bilingualism and L1/L2 acquisition is surveyed to highlight, and gain further insight into, challenges and complexities of assessing language learning in children in the FL context. Enhancing an understanding of this issue is of potentially pedagogical (as well as empirical) value as it could shed light on approaches towards establishing a framework within which to assess and interpret children's FL outcomes and progression in the way that is aligned to FL pedagogy in the primary curriculum.

It should be noted that this paper does not intend to propose that notions and evidence from bilingualism and L2 acquisition in general be directly applied to the specific Japanese EFL context. Nor does it aim to provide a comprehensive review of trends and issues in FL learning in the wider, global settings (e.g., in Europe and other oriental countries) (see Murphy, 2014 for a review). The purpose of this paper, although perhaps limited in scope, is two-fold: a) It provides a non-exhaustive overview of challenges and complexities of assessing language competence in children developing 'bilingually', b) within which the Japanese EFL context is described and context-specific issues surrounding language assessment are discussed.

Definitions and dimensions of bilingualism

Children learning more than one language could broadly be classified as bilinguals. What the term 'bilingual' exactly means is worthy of an examination here, since varied definitions and typologies of the concept *Bilingualism* have been proposed to date. They range from a native-like competence in two languages (Bloomfield, 1935) to a minimal competence in at least one of the domains in the L2 (e.g., speaking and reading) (Diebold, 1964; Macnamara, 1967). One early (and strict) definition, put forward by Bloomfield (1935), regards a learner with age-appropriate and native-like control of all aspects of two languages as a bilingual. Much broader definitions have since been proposed, which resulted in incorporating different forms of bilingualism as follows (Diebold, 1964; Tabors & Snow, 2003): a) incipient bilingualism for those who are at the beginning of acquiring some aspects of another language; b) emergent bilingualism for those with a range of (developing) abilities in the two languages; c) balanced bilingualism for whose with equal competence in the two languages; d) at-risk bilingualism for those who have become dominant in the L2, accompanied by the loss of abilities in the L1.

Another essential aspect of bilinguals' linguistic development is a distinction between receptive and productive abilities. Bilinguals with (only) receptive bilingual abilities may understand, read or interpret signs in more than one language, but are able to speak, write or produce signs only in one of the languages.

Bilinguals with (only) productive bilingual abilities possess productive skills in more than one language but receptive skills only in one language (García, 2009; Tabors & Snow, 2003). In some cases, moreover, bilinguals possess oracy abilities (i.e., listening and speaking) in one of the languages and literacy abilities (reading and writing) in the other, or in different combinations, partly related to varying levels of their proficiency across languages. Bilinguals thus typically possess varying levels of proficiency across languages particularly as they use two languages to different degrees and at different frequencies, depending on the context (school and home) and domain (speaking and writing) (Meisel, 2004). Such a varying nature of proficiency is also a consequence of social and political constraints (e.g., language practice, power, communicative needs) which exist in communities and societies (García, 2009).

Judging by the various definitions and classifications of bilinguals presented above, children learning an FL in school settings could indeed be considered one form of bilingualism, perhaps falling somewhere between incipient and emergent bilingualism. In addition, what we call monolingual children (or adults) could technically be bilinguals (or multilinguals) if they know a few words in another language, with reference to the minimalist definition of bilingualism (i.e., being able to produce/understand a few words or sentences in a language other than the L1). This minimalist definition, however, does not help us better understand the outcomes and trajectory of language learning which may be unique to each type of bilingual. Neither does it address, at least explicitly, a typology of L2 contexts, such as the difference between those learning another language in the L2 context, where the L2 is the societal (or majority) language, and those learning an FL in school settings, outside of which exposure to the target language is limited both in quality and quantity. In order to address this issue, children receiving FL instruction in school settings (whilst otherwise receiving education entirely in the L1) are termed FL learners in this paper, to make a distinction from bilingual children developing two languages through exposure to the L2 both in the home and through schooling (where the L2 is the medium of instruction for subject-matter areas). The latter can be further classified, minimally, into the following categories (Hayashi & Murphy, 2013; Murphy, 2014): simultaneous bilinguals including majority/heritage language learners, depending on which one of the two languages is in focus; and sequential bilinguals who acquired the basics of the L1 in the L1 context first, followed by learning the L2 through extensive exposure to the L2 through schooling and in the L2 community.

Assessing language development in school-age children

Language assessment is a challenging task, encompassing varied issues to be considered and overcome, which could be attributed in part to the fact that child language learners represent diverse linguistic and cultural backgrounds. The issues addressed below are intended to cover domain-general issues (and perhaps some of domain-specific (e.g., vocabulary) issues as well).

One of the greatest challenges faced by researchers (and practitioners) is to devise a 'one-size-fits-all' norm-referenced standardised measure which is appropriate for all (majority and minority) populations within a country (Cheng, 1997; Sánchez, 2006; Stockman, 2000; De Lamo White & Jin, 2011). As discussed above, there is extensive heterogeneity in language knowledge and use within each type of learner across bilingual populations, owing to the interaction of various factors, including linguistic, socio-linguistic and

socio-cultural factors. The complexity of the interplay between a child's developing language skills and his/her social and cultural environment is depicted in Bronfenbrenner's (1979) model of the ecology of human development, a theoretical framework commonly used in ethnographic studies on child language development (e.g., Cheng, 1997; Martin, 2009), as reproduced in Figure 1. Evaluating this model in detail is beyond the scope of this paper. However, a glimpse of this model, which places the child at the centre of multiple social subsystems, informs us of the interactive and complex ways in which both the direct (e.g., family and school) and indirect (e.g., educational and political systems) effects of each social subsystem shape the child's experience and language development. This model also implies that using standardised, knowledge-based assessment tools alone may lead to a limited understanding of children's developing language knowledge.

Assessing language knowledge in children is complicated also by the nature of their linguistic and metalinguistic knowledge which constantly develops as they progress with formal, extensive literacy instruction through the school years. Therefore, a genuine picture of the nature of linguistic knowledge should be obtained via the use of a valid assessment tool which not only reduces the influence of external factors on performance but also captures the rate of, and/or the direction of change in language development. Faced with these challenges, researchers have begun to employ dynamic, process-oriented measures as alternative methods to, or in addition to, static ones as exemplified by traditional knowledge-based standardised tests. Dynamic assessment attempts to measure children's potential for learning by focussing on the interaction between the assessor and the child and its influence on the performance and responsiveness of the child (Burton & Watkins, 2007; Hasson & Joffe, 2007; Peña, 2000). The theoretical grounding for this type of assessment is the notion of the zone of proximal development as defined by Vygotsky (1978: 86) as follows:

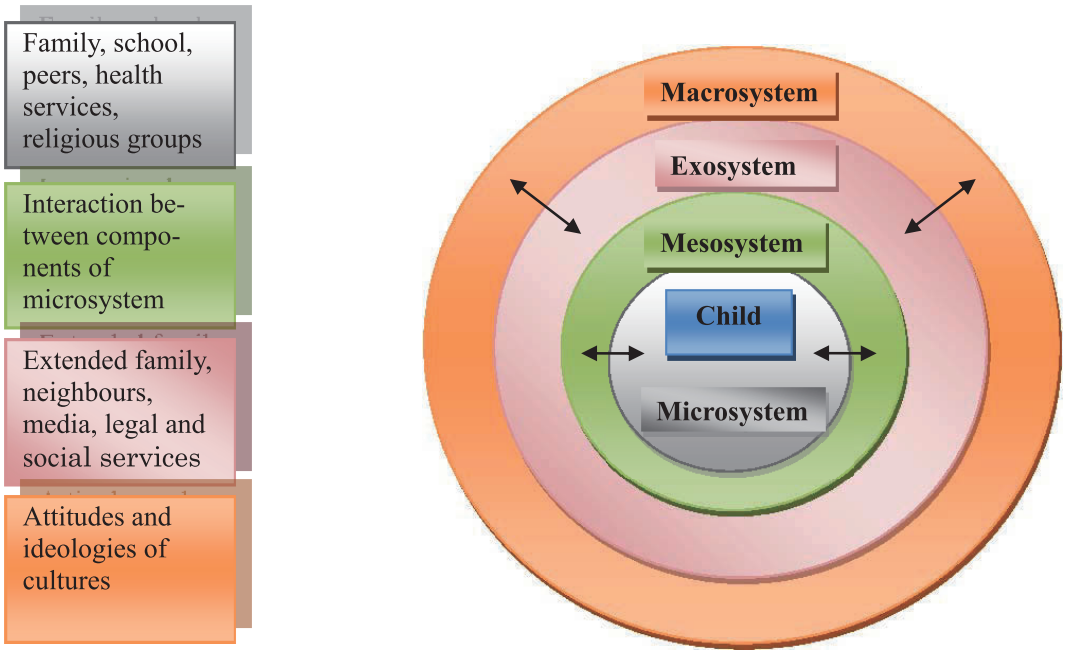

Figure 1. *Bronfenbrenner's ecology of human development*
(Bronfenbrenner, 1979, reproduced from Anderson & Van der Gaag, 2005: 91).

'the distance between the actual developmental level as determined by independent problem solving and the level of potential development as determined through problem solving under adult guidance or in collaboration with more capable peers.'

Incorporating into assessment adult-supported learning in the form of mediation or scaffolding, together with a heightened focus on the learning process, is considered to be less biased than knowledge-based standardised measures since the former minimises the role of the child's previous knowledge and experience with language (Campbell, Dollaghan, Needleman, Janosky, 1997; Peña, 2000). Other processing-dependent measures, including non-word repetition and fast-mapping tasks, are also designed to reduce such linguistic bias by tapping into degrees of efficiency with which children learn or access lexical items (Carey & Barlett, 1978; Kohnert & Kan, 2007). While they have their merits, dynamic, process-oriented assessment tools have limitations also. One issue is concerned with time constraints. Dynamic assessment is a time-consuming method, requiring the assessor to have intensive contact with the child, especially during the mediation learning phase—a phase during which the assessor or the more experienced collaborator teaches the principles of problem-solving strategies which underlie successful task completion (Peña, Iglesias & Lidz, 2001; De Lamo White & Jin, 2011). A second issue is the lack of established reliability and validity of the assessment procedures since assessors are free to modify their procedures according to the child (Hasson & Joffee, 2007). Although dynamic assessment, similar to other processing-dependent measures, helps to reduce linguistic basis by not relying solely on the child's previous experience with language, it may not perhaps be entirely free from such bias. Children with a higher language ability appear to benefit significantly more from the mediation learning experience within a dynamic assessment framework than do those with a lower language ability (Peña, et al., 2001). A third issue is the lack of qualifications as a clinician on the assessor's part. Dynamic assessment is a tool commonly used by speech-language pathologists to serve as an aid in the accurate identification and initial diagnosis of language impairments and learning difficulties in children. Extensive professional training would be required if such an assessment was to be carried out to a high standard, which might not always be readily available to the assessor(s) concerned (Hayashi, 2012).

Challenges in assessing bilingual development in young learners

As discussed thus far, assessing language competence is a highly complex task, involving the use of multiple techniques appropriate for the target sample, together with a consideration of the interaction of wide-ranging factors which may exert a significant influence on the performance of children. An assessment procedure could become even more complex when the cultural and linguistic diversity of bilingual children needs to be taken into account. Adding to this complexity is the fact that, to the researcher's knowledge, there are few to no established standardised assessment tools or procedures to date which are appropriate and also unbiased for all ethnic groups of bilingual children.

One overarching challenge in language assessment for bilingual children is to remove assessment bias, specifically, psychometric and cultural bias (Carter, Lees, Murira, Gona, Neville & Newton, 2005; Peña, 2000; Teoh, Brebner & McCormack, 2012). Psychometric bias occurs when the assessment instrument lacks ade-

quate representation of the target sample in the standardisation/norming sample and thus limits its predictive validity (Peña, 2000). The vast majority of available formal assessment tools are developed and standardised on a sample of English-speaking monolingual children (Teoh et al., 2012; Thordardottir, Rothenberg, Rivard & Naves, 2006). There is likely to be extensive variation in cultural and linguistic experiences exposed to a standardisation sample of L1 children, in comparison to those exposed to bilingual children, such as ESL (English as a second language) children in the L2 context and EFL children in Japan. In this regard, the use of monolingual-normed standardised measures for bilingual children could result in a less accurate interpretation of the results obtained and hence a less genuine profile of competence in the domain concerned. This issue has yet to be resolved satisfactorily by researchers, due to a lack of assessment tools which directly address the unique characteristics of linguistic knowledge in each type of bilingual child without reference to monolingual norms. One factor contributing to this persistent challenge is extensive heterogeneity within bilingual populations, which makes it difficult to recruit a relatively homogenous bilingual group, as required for the standardisation or validation of an assessment instrument (De Lamo White & Jin, 2011). Given the available assessment measures, recommended procedures, although perhaps not ideal, remain to be the following: to test bilingual children in both their languages using tests developed originally for the respective monolingual samples and compare their performances to those of their monolingual peers of the same age and of comparable background (Thordardottir et al., 2006). An additional justification for these procedures is the converging evidence in the literature suggesting that, while showing unique characteristics such as code-switching and language transfer (e.g., Lindholm & Padilla, 1978; Genesee, 1989), bilingual children resemble monolingual children in terms of the order of acquisition of linguistic structures (e.g., grammatical morphemes) (e.g., Dulay & Burt, 1974; Jia & Fuse, 2007; Paradis, 2005).

An additional area of assessment bias worthy of discussion is cultural bias. One example is the use of stimuli which are unfamiliar to the bilingual sample, consequently lowering their performance. Teoh et al. (2012), for instance, found that the performances of their bilingual sample in Singapore were somewhat lowered by the presence of culturally-biased pictorial stimuli from a UK-normed standardised measure (e.g., a mismatch between a target picture of a fireman in a yellow uniform with a yellow hat on and firemen in Singapore dressed in blue uniforms with red hats on). Cultural bias occurs also when the nature of the task is unfamiliar to the children. Peña and Quinn (1997), for example, showed that Puerto Rican and African American children found it difficult to name pictures in the way required by the test, owing to their previous language learning experiences which made them better prepared to describe *functions* of objects, rather than name their labels using only one word. It is also possible that not all word meanings, especially meanings attached to a polysemous word, are consistent across linguistic communities. The meaning of words, such as *foot* (of person, of bed, of mountain) and *run* (person does, water does, colour does), could vary for people from different linguistic communities due to their community-specific cultural experiences (Stockman, 2000; Yule, 2006). Standardised tests which do not tap into degrees of knowledge and hence are unable to reflect culturally different ways of word learning, could, therefore, result in underestimating lexical knowledge in bilingual children.

Another important aspect of language assessment for bilingual children is a consideration of factors beyond linguistic factors, such as psychological (e.g., religion, cultural identity) and sociological factors (e.g., language input at home, language practice in the community, parents' socioeconomic status). These factors,

to some extent, overlap with those identified as important factors for children in general, as discussed with reference to Bronfenbrenner's (1979) model (Figure 1). It is, however, reasonable to assume that the nature of language assessment is more complex with bilingual children, especially when they embrace cultural and linguistic backgrounds which are different to those represented in the monolingnual population on which assessment tools are standardised.

Challenges in assessing FL learning in young learners: the Japanese EFL context

The context in which children learn an FL in school settings is no exception to the aforementioned issues and challenges surrounding language assessment. Of particular significance is the difficulty in identifying the construct being measured. This stems from the fact that FL provision differs to a great extent from awareness-raising to language focus (or skill-based), and from content-based to immersion, perhaps both across and within contexts (Nikolove & Mihaljević Djigunović, 2011). As such, it is paramount that language assessment be carried out in the way that is closely linked to in-class activities (Murphy, 2014). In the Japanese EFL context, as mentioned above, English language learning as a compulsory area of study, known as FLA, was introduced in 2008 and has been fully implemented to pupils in Years 5 and 6 in primary schools in Japan since 2011. Since it is not regarded as a subject, no formal, numerical evaluation is to be carried out. As such, children's outcomes and progression in the learning of English are to be assessed using descriptive measures, such as a collection of reflection sheets filled out by pupils, observation notes made by the teacher and so on. An overall evaluation by the teacher is expressed in sentential format similar to, or modelled on, can-do statements (see below), as opposed to numeral ratings (e.g., on a 3-point scale). Further, no formal literacy instruction (i.e., reading and writing) is given as yet, since the primary objectives of FLA lie in familiarising pupils with (English) sounds and basic expressions through communicative activities which involve the use of oracy abilities (MEXT, 2008). Such an approach is designed to serve as a basis for the acquisition of communication skills of a more comprehensive nature inclusive of all four domains (i.e., listening, reading, writing and speaking) (Higuchi, 2013).

Assessing oracy abilities and vocabulary knowledge

Emergent in the literature is a body of evidence of English language abilities, especially listening and speaking abilities and vocabulary knowledge, in Japanese pupils of primary school age, as measured by experimental and/or standardised tests. A large-scale study was carried out by Goto Butler and Takeuchi (2006), who measured the communication abilities of primary pupils (Years 1–6) in 28 schools across Japan, using a standardised English listening test called *Bronze*. This test constitutes a suite of subtests called *Jido -Eiken* (http://www.eiken.or.jp/jr_step/), which comes in three versions—*Bronze*, *Silver* and *Gold*—in the ascending order of difficulty. Goto Butler and Takeuchi showed that the total number of hours spent on learning English in school was positively related to gains in listening ability. As cautioned by the researchers themselves, however, this finding may not be generalisable to the wider primary school context in Japan since the participants had been recruited from volunteering schools which had already been proactive in English language provision beforehand. These schools, at the time of this study, accounted for approximately 20 percent only of primary schools in Japan.

Another large-scale study carried out by National Institute for Education policy Research (NIER, 2009) was aimed at establishing profiles of English language (oracy) abilities in primary school pupils, in response to the nationwide introduction of FLA. Listening and speaking tests were administered to Year 5 and 6 groups of approximately 3,300 and 400 pupils, respectively. The speaking test was administered individually in 5-minute face-to-face interview format, which entailed a range of short question-answer interactions with pictures used as an aid for some of the questions. The test was carried out on topics presumed to be familiar to the pupils such as greetings, colours, animals, food, sport, etc. Whilst accuracy rates in oral responses, overall, increased for pupils in Year 6, both the Year 5 and Year 6 groups appear to have had difficulty responding to questions concerning numbers (e.g., 'How much is this bag?', 'It's 800 yen') and subjects (e.g., 'Do you like math?', 'Yes, I do / No, I don't.'). The listening test consists of five subcategories—vocabulary (especially, nouns and verbs), classroom English (e.g., 'Stand up, please'), short dialogues, question-response, and comprehension of short (spoken) text. The results of this test showed that the pupils performed accurately on the vocabulary, classroom English and short dialogue sections. Accuracy rates, in contrast, lowered in the other sections of a more cognitively demanding nature, where the children had to process multiple sources of input consisting of pictorial and sentential stimuli in choosing the appropriate response (e.g., 'Where do you want to go?' in the question-response section), and to match a multi-sentential stimulus (e.g., 'I'm in the classroom. I have numbers. I have hands.' in the short-text comprehension section) to the picture given. Similar findings were obtained in other recent studies, where children scored low on sections involving top-down processing in listening (e.g., comprehending the content of multi-sentential text describing situations and dialogues) (Ishihama, Watanabe & Someya, 2014) and also in speaking (e.g., summarising a story) (Yukawa, Takanashi & Koyama, 2008), relative to sections involving bottom-up processing (e.g., word-picture matching).

A study by Sakuma (2011) illuminates children's listening ability further from a cognitive perspective by investigating the relationship between listening ability and working memory capacity in primary pupils in Years 1–6 in Japan. High scores on the *Bronze* version of *Jido-Eiken* were associated with accurate performance on Japanese (Years 1, 4, 5) or English (Year 2) working memory (digit span) tasks, thus suggesting the potential efficacy of employing different memory strategies in listening activities across school years. Such an association was not identified in Year 3 or 6 pupils, however. Sakuma explained this by suggesting that there might have been ceiling effects of the Bronze test for Year 6 pupils, whilst also highlighting difficulty in identifying cognitive factors underlying the developing cognitive abilities of children (Year 3) transitioning to the latter part of primary school education. This finding, in turn, leaves open the question of whether or not the test is sensitive enough to age-related differences in listening ability.

An increasing body of evidence has been gathered of vocabulary knowledge in children in the Japanese EFL context, ranging from studies on incidental and intentional vocabulary learning (e.g., Karreira-Matsuzaki, Shigyo, Simoda & Sakamoto, 2010; Nakamura, Suematsu & Hayashida, 2010; Yoshimura, 2009) to studies on the development of a corpus-based vocabulary list for EFL learners in primary and secondary school settings (e.g., Hasegawa & Machida, 2010; Ishikawa, 2005; Nishigaki, Chujo & Kashimura, 2007). Karreira et al. (2010), for instance, investigated the effects of ICT reading materials called LeapFrog TAG Reading System, whereby clicking on each picture produces music and other sound effects, on gains in vocabulary knowledge in 70 pupils in Year 5. The children were tested on 10 target items before and after

the 10-day reading programme. Whilst the children's attitude towards FL (English) reading did not change significantly across testing points, they scored higher on six of the target items at post-test, thus indicating possible vocabulary gains as a result of reading. This finding should be interpreted with caution, however, due to a few methodological limitations. As also raised by the authors, there was no control group involved. An additional issue is that a response to each target item was elicited via a Japanese question (e.g., 'Eigode *katatsumuri* ga iemasuka (Can you say 'snails' in English?)'), followed by a yes/no response in Japanese. That is, only receptive aspects (recognising and recalling) of the connection between the target word and its definition were measured. Further, the books used for the programme were originally designed for English-L1 children in English-speaking environments. Whilst such a bespoke vocabulary test may be closely aligned to the activities carried out, these issues in turn call into question its predictive validity, as raised also in the section on language assessment for bilingual children above. Having said that, the status quo of English being learnt as an area of study, not a subject, makes it difficult to resolve this matter, due to the lack of official attainment levels established and also to the lack of standardised tests of language competence for young L2/FL learners in general (Murphy, 2014).

Common European Framework of Reference Japan (CEFR-J)

One language framework which has increasingly been used for FL assessment as well as FL pedagogy and curriculum development across Japan is the Common European Framework of Reference Japan (CEFR-J) (Negishi, 2011). CEFR-J is modelled on the Common European Framework of Reference for Languages (CEFR), which 'describes in a comprehensive way what language learners have to learn to do in order to use a language for communication and what knowledge and skills they have to develop so as to be able to act effectively' (Council of Europe, 2001: 1). The framework provides a set of proficiency statements known as 'can-do statements (or descriptors)' illustrating what learners are capable of using the language concerned at any given point in time (North, 2007). The can-do statements are classified into the following six proficiency levels under three categories, in relation to five skills (listening, reading, spoken interaction, spoken production, writing): Basic User (A1 and A2); Independent User (B1 and B2); and Proficient User (C1 and C2). It should be noted that providing a comprehensive of review of both pros and cons for the application of CEFR or the CEFR-J itself, is beyond the scope of this paper. A brief overview of relevant issues is provided below, to add further insight into the challenges and issues surrounding language assessment in the Japanese EFL context.

The CEFR-J was introduced to the Japanese context, having addressed an urgent need for standards which would reflect and fit the needs and situations of FL education (English, in particular) in Japan (Negishi, 2011). A level preceding the A1 level, labelled Pre-A1, was created, and the A1 level was further classified into three sub-levels (Tono & Negishi, 2012): A1.1, A1.2 and A1.3. The A2 to B2 levels are each further divided into two sub-levels: A2.1, A2.2, B1.1, B1.2, B2.1, B2.2. Such a finer classification in proficiency was motivated by the fact that the vast majority of Japanese FL learners (approximately 80 percent) are skewed towards the A level of the CEFR scale (Negishi, 2011). The CEFR-J has been increasingly applied to the Japanese educational context, acting primarily as a tool for providing Japanese proficiency standards for foreign languages, interpreting high-stake proficiency test scores (e.g., TOEIC, TOEFL), enhancing transparency and coherence in educational provision (e.g., curricular choices and examinations), and also

for empowering autonomous language learning (CoE, 2001; Nagai & O'Dwyer, 2011).

Whilst increasingly researchers argue for the benefits of applying the CEFR to the Japanese EFL context (or FL pedagogy in general) (e.g., Naganuma, 2011; Takahashi & Yanagi, 2009), it is without limitations. One limitation is the lack of clarity as to how mastery is defined, especially in terms of how likely it is for a learner at a certain level to succeed in carrying out the can-do statements, as it is rather unrealistic to assume that the learner will always perform perfectly on the task concerned (Runnels, 2013). Further, whilst these statements are informative concerning what learners at a certain level can do, little is known about what they should know (e.g., specific grammar points), in order to carry out the linguistic tasks specified in the statements (Westhoff, 2007). Relevant to this limitation is the lack of empirical research—30 years of research into second language acquisition (SLA)—underpinning the CEFR, which renders it difficult to achieve an enhanced understanding of how L2 or FL proficiency develops across levels (Alderson, 2007; Weir, 2005). As a result, for instance, it remains as yet unclear what theory of comprehension should be used to identify the mental operations underlying the performance of learners at different levels, thereby leaving little room for a diagnosis or interpretation of the development of receptive abilities in the CERF terms (Alderson, 2007). More recently, a study by Runnels (2013) provides evidence which raises concerns for the sensitivity of the sublevels posited by the CEFR-J. A survey was carried out, where 590 university EFL learners rated, on a 5-point Likert scale, the extent of their agreement with all 50 Japanese can-do statements for all five skills from levels A1.1 to A2.2. The results indicated that no significant differences in rating were identified between the adjacent sublevels of the A1 and A2 levels, leading Runnels to suggest that the finer distinction in proficiency advanced by the CEFR-J may not be as realistic as it is purported to be.

To reiterate, accepting or refuting the efficacy of the CEFR or its application to the Japanese context is beyond the scope of this paper. The discussion presented thus far serves to provide a brief overview of the more context-specific standards for FL proficiency, referred to as the CEFR-J, together with issues and challenges associated with its implementation on a wider scale. Since the CEFR is framed in a (deliberately) language-independent manner, much more empirical research is needed to illuminate the trajectory of L2/FL learning—how proficiency develops over time at a more local level (i.e., in a particular L2/FL) (Alderson, 2007) . Further, more empirical projects evaluating the efficacy of the implementation of the CEFR are warranted in order for the use of the framework to 'bring curricula, pedagogy, and assessment into fruitful interaction with one another' (Little, 2007: 652).

Concluding remarks

Assessing the outcomes and progression of language learning in young learners is a considerable challenge. In order to obtain a comprehensive picture of the nature of language competence in children learning more than one language, it is perhaps recommended that researchers employ multiple techniques, such as the use of static, knowledge-based measures, in conjunction with more dynamic, processing-dependent measures. Longitudinal investigations, especially within an ethnographic design which takes multiple sociocultural factors (see Figure 1) into consideration, would not only add more detail to our understanding of children's developing FL/L2 knowledge but also help to enhance the reliability and validity of the data col-

lected. Such approaches, although ideal, may not always be feasible due to a constellation of factors (e.g., educational, cultural and practical). In Japanese primary school settings, establishing a common framework for assessing FL competence remains a challenge, particularly as assessment is typically tied to in-class activities within the primary curriculum and also there persists a paucity of available standardised tests. One way forward is the introduction of the CEFR-J and future empirical investigations should shed light on its efficacy in the wider Japanese FL context.

Acknowledgements

I am very grateful to Victoria A. Murphy for her valuable input and feedback especially on the bilingual/SLA aspects of the paper. Special thanks also extend to Shoichi Tanaka and Matthew C. Clarke for valuable comments and discussions at stages leading up to the final draft of the paper.

References

Alderson, J. C. (2007). The CEFR and the need for more research. *The Modern Language Journal, 91*(4), 659-663.

Anderson, C., & Van Der Gaag, A. (2005). *Speech and language therapy: Issues in professional practice*. London: Whurr.

Bloomfield, L. (1935). *Language*. London: Allen & Unwin.

Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Cambridge, MA: Harvard University Press.

Brown, R. (1973). *First language*. Cambridge MA: Harvard University Press.

Burton, V. J., & Watkins, R. V. (2007). Measuring word learning: Dynamic versus static assessment of kindergarten vocabulary. *Journal of Communication Disorders, 40*(5), 335-356.

Campbell, T., Dollaghan, C., Needleman, H., & Janosky, J. (1997). Reducing bias in language assessment: Processing-dependent measures. *Journal of Speech, Language and Hearing Research, 40*(3), 519-525.

Carey, S., & Bartlett, E. (1978). Acquiring a single new word. *Papers and Reports on Child Language Development, 15*, 17-29.

Carter, J. A., Lees, J. A., Murira, G. M., Gona, J., Neville, B. G. R., & Newton, C. R. J. C. (2005). Issues in the development of cross-cultural assessments of speech and language for children. *International Journal of Language & Communication Disorders, 40*(4), 385-401.

Cheng, L. R. L. (1997). Diversity: Challenges and implications for assessment. *Journal of Children's Communication Development, 19*(1), 55-62.

Council of Europe. (2001). *The Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

De Lamo White, C., & Jin, L. (2011). Evaluation of speech and language assessment approaches with bilingual children. *International Journal of Language & Communication Disorders, 46*(6), 613-627.

Diebold, A. R. (1964). *Language in culture and society*. New York: Harper Collins.

Dulay, H. C., & Burt, M. K. (1974). Natural sequences in child second language acquisition. *Language Learning, 24*(1), 37-53.

García, O. (2009). *Bilingual education in the 21st century: A global perspective*. Malden, MA ; Oxford: Wiley-Blackwell.

Genesee, F. (1989). Early bilingual development: One language or two. *Journal of Child Language, 16*(1), 161-179.

Goto Butler, Y., & Takeuchi, A. (2006). Shogakkou eigokatsudou ni okeru hyoka—Jido-Eiken wo tsukatta kokoromi [Assessing English language learning in foreign language activities using *Jido-Eiken* (Bronze)]. *Japan Association for the Study of Teaching English to Children (JASTEC)*, Vol.25, 1-16.

Gregory, E. (2008). *Learning to read in a new language: Making sense of words and worlds* (2nd ed.). London: SAGE Publications.

Hasegawa, N., & Machida, H. (2010). Kodomo no gengoryoku wo saguru: Eigo goi tesuto to nihongo goi nouryoku tesuto no kekka kara [Langauge competence in children: Results from tests of English and Japanese vocabulary]. *Scientific Approaches to Language, 9*, 191-213.

Hasson, N., & Joffe, V. (2007). The case for dynamic assessment in speech and language therapy. *Child Language Teaching*

and Therapy, 23(1), 9-25.

Hayashi, Y. (2012). *On the nature of morphological awareness and vocabulary knowledge in school-age English-Japanese monolingual and bilingual children*. Unpublished Ph.D., University of Oxford, Oxford.

Hayashi, Y., & Murphy, V. A. (2013). On the nature of morphological awareness in Japanese-English bilingual children: A cross -linguistic perspective. *Bilingualism: Language and Cognition, 16*(01), 49-67.

Higuchi, T. (2013). *Shogakko eigokyoikuhou nyūmon* [*An introduction to English education in elementary school*]. Kenkyūsha.

Ishihama, H., Watanabe, T., & Someya, F. (2014). 'Hi friends! 1' ni junkyoshita chokairyoku tesuto no kaihatsu to sono ouyoukekka ni kansuru jireihoukoku (2) [A report on the development of a listening comprehension test based on 'hi friends! 1' and its applicability]. *The 14th Japanese Association of English Teaching in Elementary Schools Annual Conference*, Kanto Gakuin University, Kanagawa. pp. 1-12.

Ishikawa, S. (2005). Nihonjin jidouyou eigo kihongoihyou kaihatsu ni okeru hindo to ninchido no mondai: Bogo corpus to taishougo corpus no hindo yugou no shuhou [Frequency and familiarity in compiling the English word list for children— Based on the two corpora of the mother tongue and the target language. *Technical Report of IEICE. Thought and Language, 105*(437), 43-48.

Jia, G., & Fuse, A. (2007). Acquisition of English grammatical morphology by native Mandarin-speaking children and adolescents: Age-related differences. *Journal of Speech, Language, and Hearing Research, 50*(5), 1280-1299.

Karreira-Matsuzaki, J., Shigyo, T., & Shimoda, Y. (2010). Otono deru eigo no ehon tag reading system wo toriireta shogakkou gaikokugokatsudou [EFL classes incorporating tag reading system in a Japanese elementary school]. *Computer & Education, 29*, 12-17.

Kohnert, K., & Kan, P. (2007). Lexical skills in young children learning a second language: Methods, results, and clinical applications. In J. G. Centeno, R. T. Anderson & L. K. Obler (Eds.), *Communication disorders in Spanish speakers: Theoretical, research and clinical aspects* (pp. 156-168). Clevedon: Multilingual Matters.

Lindholm, K. J., & Padilla, A. M. (1978). Language mixing in bilingual children. *Journal of Child Language, 5*(2), 327-335.

Little, D. (2007). The common European framework of reference for languages: Perspectives on the making of supranational language education policy. *The Modern Language Journal, 91*(4), 645-655.

Macnamara, J. (1967). The bilingual's linguistic performance—a psychological overview. *Journal of Social Issues, 23*(2), 58-77.

Martin, D. (2009). *Language disabilities in cultural and linguistic diversity*. Bristol: Multilingual Matters.

Meisel, J. M. (2004). The bilingual child. In T. K. Bhatia, & W. C. Ritchie (Eds.), (pp. 91-113). Oxford: Wiley-Blackwell.

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2008). *Shin gakushūshidoyōryō [Course of Study for primary schools.]* Retrieved February 5, 2012, from http://www.mext.go.jp/a_menu/shotou/new-cs/youryou/syo/sou.htm.

Ministry of Education, Culture, Sports, Science and Technology (MEXT). (2014). *Kongo no eigokyoiku no kaizen jujitsu housaku ni tsuite houkoku —global ka ni taioushita eigokyoiku kaikaku no itsutsu no teigen [A report on plans for improving English language education in Japan: 5 proposals in the era of globalisation]*. Retrieved 28th October, 2014, from http://www.mext.go.jp/b_menu/shingi/chousa/shotou/102/houkoku/1352460.htm.

Murphy, V. A. (2014). *Second language learning in the early school years: Trends and contexts*. Oxford: Oxford University Press.

Nagai, N., & O'Dwyer, F. (2011). The actual and potential impacts of the CEFR on language education in japan. *Synergies Europe, 6*, 141-152.

Naganuma, N. (2011). Shogakkou eigokatsudo ni okeru jiritusei to doukizuke wo takameru can-do hyoka no jissen [the implementation of can-do assessment to enhance autonomy and motivation in the elementary school English activity]. *ARCLE REVIEW, 5*, 65-74.

Nakamura, N., Suematsu, A., & Hayashida, H. (2010). Shogakkou eigo ga chugakkou no eigogakushu ni oyobosu eikyo ni tsuite — goi no jikohyouka ni shouten wo atete [Effects of English activities in elementary schools on English language learning in junior high schools — self-evaluation of vocabulary learning in focus]. *JES Bulletin, 10*, 25-30.

National Institute for Educational Policy and Research. (2009). *Heisei 20nendo 'shogakko ni okeru eigokyoiku no arikata ni kansuru chosakenkyu' seika houkokusho [A report on foreign language education in primary schools for the 2008 academic year*. Retrieved 26th November, 2014, from http://www.nier.go.jp/shoei_h20/shoei.html.

Negishi, M. (2011). CEFR-J Kaihatsu no Keii [The Development Process of the CEFR-J]. *ARCLE Review, 5*(3), 37-52.

Nikolov, M., & Mihaljević Djigunović, J. (2011). All shades of every color: An overview of early teaching and learning of foreign languages. *Annual Review of Applied Linguistics, 31*, 95-119.

Nishigaki, C., Chujo, K., & Kashimura, M. (2007). Shogakkoueigo ni okeru nichijou seikatsu goi no shidou [teaching daily life English vocabulary to Japanese elementary school children—creation and use of 'Karuta' cards—. *Bulletin of the Faculty of Education, Chiba University, 55*, 255-270.

North, B. (2007). The CEFR Common Reference Levels: Validated reference points and local strategies. *Language Policy Forum Report*, 19-29.

Paradis, J. (2005). Grammatical morphology in children learning English as a second language implications of similarities with specific language impairment. *Language, Speech, and Hearing Services in Schools, 36*(3), 172-187.

Peña, E. D. (2000). Measurement of modifiability in children from culturally and linguistically diverse backgrounds. *Communication Disorders Quarterly, 21*(2), 87-97.

Peña, E. D., Iglesias, A., & Lidz, C. S. (2001). Reducing test bias through dynamic assessment of children's word learning ability. *American Journal of Speech-Language Pathology, 10*(2), 138-154.

Peña, E. D., & Quinn, R. (1997). Task familiarity: Effects on the test performance of Puerto Rican and African American children. *Language, Speech, and Hearing Services in Schools, 28*(4), 323-332.

Runnels, J. (2013). Preliminary validation of the A1 and A2 sub-levels of the CEFR-J. *Shiken Research Bulletin, 17*(1), 3-10.

Sakuma, Y. (2011). Cognitive features of working memory in elementary school students participating in foreign language activities. *ARELE: Annual Review of English Language Education in Japan, 22*, 233-248.

Sánchez, L. (2006). Bilingualism/second-language research and the assessment of oral proficiency in minority bilingual children. *Language Assessment Quarterly: An International Journal, 3*(2), 117-149.

Stockman, I. J. (2000). The new Peabody Picture Vocabulary Test-III: An illusion of unbiased assessment? *Language, Speech, and Hearing Services in Schools, 31*(4), 340-353.

Tabors, P. O., & Snow, C. E. (2003). Young bilingual children and early literacy development In S. B. Neuman, & D. K. Dickinson (Eds.), *Handbook of early literacy research* (pp. 159-178). New York; London: Guildford Press.

Takahashi, M., & Yanagi, Y. (2009). Taikenteki na gengo gakushu ni yoru mojishidou ni okeru touitsu hyoukakijun no kenkyu: Europe kyotsusanshowaku (CEFR) wo moto ni shite [A study of evaluation standards for the learning of letters in foreign language activities, based on the CEFR]. *JES Bulletin*, (9), 103-110.

Teoh, W. Q., Brebner, C., & McCormack, P. (2012). Assessing the language skills of children from culturally and linguistically diverse backgrounds: The expressive vocabulary performance of Singaporean English-Mandarin bilingual pre-schoolers. *International Journal of Speech-Language Pathology, 14*(3), 281-291.

Thordardottir, E., Rothenberg, A., Rivard, M. E., & Naves, R. (2006). Bilingual assessment: Can overall proficiency be estimated from separate measurement of two languages? *Journal of Multilingual Communication Disorders, 4*(1), 1-21.

Tono, Y., & Negishi, M. (2012). The CEFR-J: Adapting the CEFR for English language teaching in Japan. *Framework & Language Portfolio SIG Newsletter, 8*, 5-12.

Vygotsky, L. S. (1978). *Mind and society: The development of higher mental processes.* Cambridge, MA: Harvard University Press.

Weir, C. J. (2005). Limitations of the common European framework for developing comparable examinations and tests. *Language Testing, 22*(3), 281-300.

Westhoff, G. (2007). Challenges and opportunities of the CEFR for reimagining foreign language pedagogy. *The Modern Language Journal, 91*(4), 676-679.

Yoshimura, H. (2009). Kouritsu shogakkou jidou no goichishiki ni kansuru chosa no jissenhoukoku [A report on the nature of vocabulary knowledge in primary school children at state schools]. *JES Bulletin, 9*, 23-30.

Yukawa, E., Takanashi, T., & Koyama, T. (2008). Nihonjinshogakusei no eigo communication noryoku [Japanese elementary school students' communicative competence in English]. *Mother Tongue, Heritage Language and Bilingual Education Research, 4*, 68-85.

Yule, G. (2006). *The study of language* (3rd ed.). Cambridge, UK: Cambridge University Press.